Resource

Analysis of the Human Endogenous Coregulator Complexome

Anna Malovannaya,[1,2,4] Rainer B. Lanz,[1,4] Sung Yun Jung,[1,2] Yaroslava Bulynko,[1] Nguyen T. Le,[2] Doug W. Chan,[1,2] Chen Ding,[2] Yi Shi,[2] Nur Yucer,[2] Giedre Krenciute,[2] Beom-Jun Kim,[2] Chunshu Li,[2] Rui Chen,[3] Wei Li,[1] Yi Wang,[1,2] Bert W. O'Malley,[1,5,*] and Jun Qin[1,2,5,*]

[1]Department of Molecular and Cellular Biology
[2]Center for Molecular Discovery, Verna and Marrs McLean Department of Biochemistry and Molecular Biology
[3]Department of Molecular and Human Genetics
Baylor College of Medicine, Houston, TX 77030, USA
[4]These authors contributed equally to this work
[5]These authors contributed equally to this work
*Correspondence: berto@bcm.edu (B.W.O.), jqin@bcm.edu (J.Q.)
DOI 10.1016/j.cell.2011.05.006

SUMMARY

Elucidation of endogenous cellular protein-protein interactions and their networks is most desirable for biological studies. Here we report our study of endogenous human coregulator protein complex networks obtained from integrative mass spectrometry-based analysis of 3290 affinity purifications. By preserving weak protein interactions during complex isolation and utilizing high levels of reciprocity in the large dataset, we identified many unreported protein associations, such as a transcriptional network formed by ZMYND8, ZNF687, and ZNF592. Furthermore, our work revealed a tiered interplay within networks that share common proteins, providing a conceptual organization of a cellular proteome composed of minimal endogenous modules (MEMOs), complex isoforms (uniCOREs), and regulatory complex-complex interaction networks (CCIs). This resource will effectively fill a void in linking correlative genomic studies with an understanding of transcriptional regulatory protein functions within the proteome for formulation and testing of future hypotheses.

INTRODUCTION

Protein-protein interactions constitute the molecular backbone of cell biology, where select proteins assemble into metastable complexes to form bioactive units (Alberts, 1998; Köcher and Superti-Furga, 2007). These complexes then dynamically associate with each other in context of larger networks to carry out diverse biological functions. Thus, understanding the basic mechanisms of cell homeostasis requires knowledge of both the composition of protein complexes and the interactions between them.

A systems biology view of protein interactions has begun to emerge from large-scale studies in model organisms such as yeast, worms, and fruit flies (Gavin et al., 2006; Giot et al., 2003; Ito et al., 2001; Krogan et al., 2006; Li et al., 2004; Uetz et al., 2000). These analyses were made possible due to the development of high-throughput (HT) methods for measuring protein-protein interactions by affinity purification of tagged protein baits followed by mass spectrometry (AP/MS) and yeast two-hybrid assays. Limitations in genetic manipulations hinder such studies in human cells. We thus developed a protocol for HT isolation and identification of endogenous protein complexes from human cell lines using primary antibody immunoprecipitation and mass spectrometry (IP/MS). We also addressed key limitations associated with such studies, which are cross-reactivity of primary antibodies and nonspecific binding, and proposed an approach for the deconvolution of HT-IP/MS data into discrete protein complexes (Malovannaya et al., 2010).

Our current work was inspired by—and for a large part included—the Nuclear Receptor Signaling Atlas (NURSA; http://www.nursa.org) consortium proteomics effort, whose goal is to systematically isolate and identify the human nuclear receptor (NR) coregulator complexome. NR coregulators are a diverse group of molecules that associate with sequence-specific transcription factors to collectively modulate target gene expression (Lonard and O'Malley, 2007; Lonard et al., 2007; McKenna et al., 1999; O'Malley et al., 2008; Weake and Workman, 2010). Initial biochemical isolations of the mammalian coactivators Mediator and BAF/P-BAF and HDAC corepressors revealed that many coregulators assemble into multisubunit protein complexes (Gu et al., 1999; Guenther et al., 2000; Wang et al., 1996; Xue et al., 1998), implying that a comprehensive picture of the protein interaction networks is needed to better understand the regulation of biological processes in the cell.

This study presents the most extensive interaction dataset for endogenous regulatory human proteins obtained to date. By preserving both stable and weak protein interactions during complex isolations, we unveiled a modular and hierarchical organization of protein complex networks that serve as a blueprint for a better understanding of mechanisms of mammalian cell regulation. Our approach can be used in other biological systems as well; with broad applicability in mind we therefore discuss the analysis schema we used for the definition and annotation of protein complexes. Knowing the composition of protein complexes and
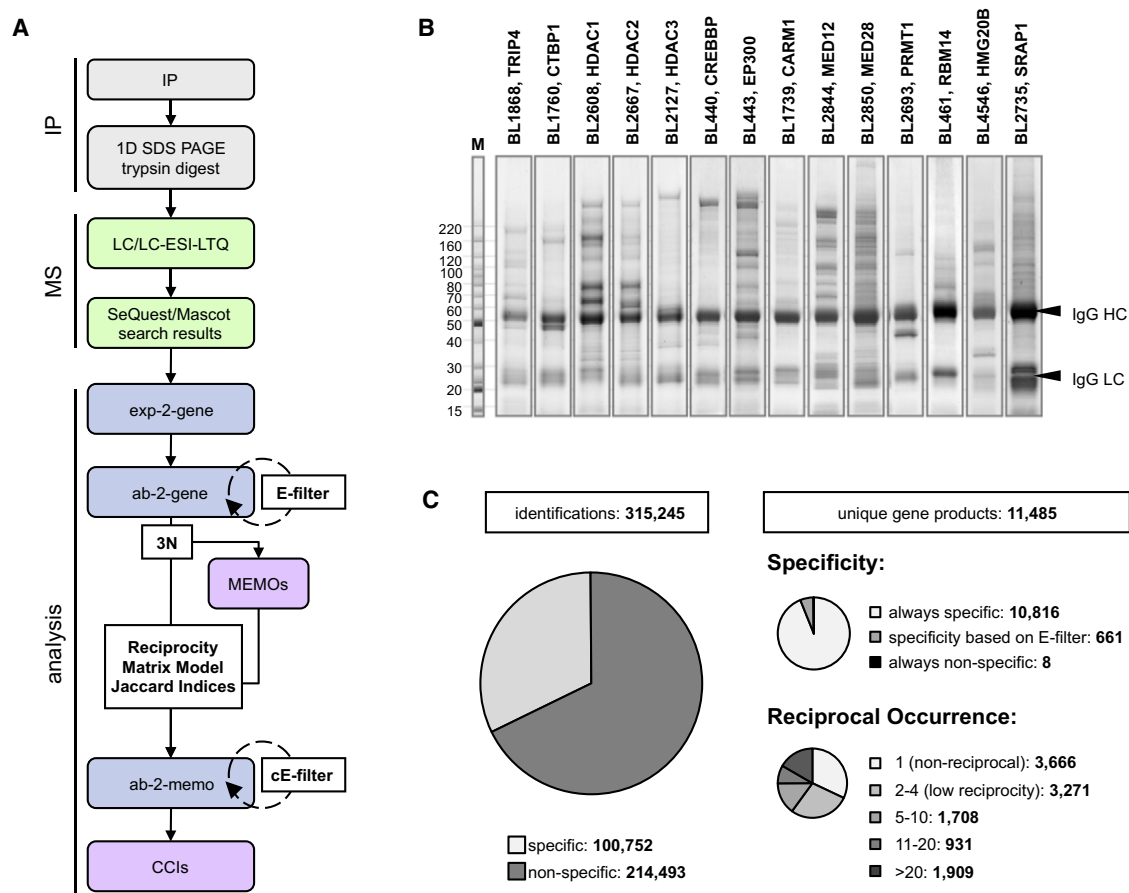
Figure 1. HT-IP/MS Analysis of the Human Endogenous Complexome

(A) Our workflow consists of IP/MS followed by filtering of nonspecific identifications, definition of minimal core complex modules (MEMOs), and assignment of complex-complex interactions (CCIs). Data relationships are abbreviated as "exp-2-gene," "ab-2-gene," and "ab-2-memo" for experiment-to-genes, antibody-to-genes, and antibody-2-MEMOs, respectively.

(B) Representative SDS-PAGE of NR coregulator complexes. IgG HC and LC are heavy and light chains of primary antibodies.

(C) Approximately 40% of the human gene products were recovered in our HT-IP/MS data in ~100,000 specific protein identifications. Majority of these were found in at least two or more antibody-different experiments, laying a foundation for finding reciprocally verified protein associations (additional descriptive statistics and data are available in Figure S1 and Table S1).

their interaction networks ultimately will allow for an effective translation of genomic data to functional molecular biology. Such resources, therefore, have the potential to transform the field of basic research and greatly impact translational efforts.

RESULTS AND DISCUSSION

A Comprehensive HT-IP/MS of the Human Endogenous Complexome

To initiate delineation of the "regulatory" human protein complexome, we preferentially targeted transcriptional and signaling proteins. We used 1796 primary antibodies in 3290 immunoprecipitation experiments (Figure 1 and Table S1 and Table S2 available online) to obtain a total of ~300,000 protein identifications, about 100,000 of which we deemed specific after in silico data filtering (Figures 1A and 1C). These identifications represent 11,485 unique human gene products, more than half of which we recovered redundantly by using different antibodies.

To assess the overall discovery potential, we compared all possible unique pairs of protein associations from our dataset with human protein interactions found in the CORUM, BioGRID, IntAct, and HPRD resources (Figure S1). Our dataset covers about 70% of the protein associations reported in CORUM (Ruepp et al., 2010), whereas most of our high-confidence interactions are not listed in this nor any other current public repositories. Based on this comparison alone, our single-source study recovered the majority of protein complexes previously reported in the literature but, at the same time, significantly expanded the human interactome.

Deconvolution of the IP/MS Data into Organized Cellular Complexome

The nature of IP/MS data output—a list of protein identifications without indications of specific interactions within and between complexes—commands a different logic for data analysis as compared to conventional protein-protein interaction (PPI)

mapping. We thus developed an analysis schema that reveals an intrinsic tiered organization of the interactome in three discrete layers. These are (1) the *m*inimal *e*ndogenous core complex *mo*dules ("MEMOs"), (2) the *uni*que *core* complex isoforms ("uniCOREs"), and (3) the *c*omplex-*c*omplex *i*nteraction networks (CCIs) (Figure S2A). Such classification respects protein complex modularity and heterogeneity, which constitute a fundamental program of cellular organization, and aids in delineation of higher-order protein interactions on a proteome-wide level.

Minimal Endogenous Modules and Core Complex Isoforms

A core protein complex is loosely defined in biochemistry as strongly associated proteins that resist separation in column fractionations. Many core complexes, however, exhibit some level of variability in only a few components of otherwise similar protein assemblies. Such complex "isoforms" can be isolated and shown to function exclusively of each other. For example, we recovered the MSL and NSL complexes that contain transcriptional regulator MYST1 histone acetyltransferase and recently have been shown to display different substrate specificities (Cai et al., 2010; Mendjan et al., 2006). To better convey such modularity, we identified variable components in protein complexes and formed a dataset of MEMOs. These modules represent invariant minimal complexes of proteins with stoichiometric interdependence across the entire IP/MS dataset and serve as the conceptual building blocks of the protein complexome. Ultimately, each of the human protein-coding gene products will be assigned to one MEMO only. MEMOs are then used to reconstitute all distinct protein complex isoforms, which we call uniCOREs. The multisubunit protein complexes conventionally described in the literature most closely correspond to uniCOREs in our resource and likely impart biological functional classification.

We recently have shown that high-confidence protein interaction information can be derived from systematic IP/MS studies by enforcing two key constraints: reciprocal coappearance of proteins in multiple antibody-different IPs and preservation of stoichiometry of the core components (Malovannaya et al., 2010). Operationally, MEMOs are identified by running reciprocal near neighbor network (3N) analyses iteratively over the entire dataset to find sets of proteins with the highest correlations in different experiments. Proteins that do not show consistency in stoichiometric interactions form "singletons," which are MEMOs with special implications, discussed later using SPEN and SRC-3 as examples.

De Novo Derivation of Complex-Complex Interactions

Although much insight is gained from the core protein complexes, a more daunting task is to acquire information of the CCIs. In fact, CCIs may represent the backbone of regulatory biology; yet, on no account have they been systematically addressed. We show here that CCIs can be derived de novo from HT-IP/MS data.

Transient interactions between protein complexes do not show a stoichiometric dependence; we therefore used Boolean metrics on co-occurrences of MEMOs to find complex-complex interactions. Briefly, we first calculated reciprocity and a series of

Jaccard indices for all protein associations (see Extended Experimental Procedures). Furthermore, we adopted a matrix model score previously described as part of socio-affinity index (Gavin et al., 2006) to capture above-chance associations and defined a CCI rank that combines protein interaction scores for all subunits of a given MEMO as a chance indicator for true associations between MEMOs (Extended Experimental Procedures). We have precalculated all CCIs for each gene product in our HT-IP/MS dataset and formed a resource of human endogenous MEMOs and CCI networks that allows researchers to interrogate our data (see Figure S6). Subsequent iterative searching for different MEMO components can be used to identify uniCOREs, as described next.

Example of Assignment of uniCOREs through Reciprocal Exclusion in CCI Networks: The BRCA1 Extended Interaction Network

BRCA1 (*br*east *ca*ncer 1) has been extensively studied in DNA-damage responses as well as transcriptional regulation. Multiple IP/MS experiments showed that the BRCA1 interactome is more heterogeneous and modular than previously thought. BRCA1 and BARD1 (*B*RCA1 *a*ssociated *R*ING *d*omain 1) show stoichiometric interdependency and constitute a MEMO, and they are found in a pool of apparently stable complexes either with *BR*CA1-*i*nteracting *p*rotein 1 (BRIP1), PALB2/BRCA2, RBBP8, BRE/BRCC3/C19orf62, FAM175A (Abraxas), FAM175B (Abraxas brother 1), or UIMC1 (Figure 2A and Figure S2). These proteins do not have consistent stoichiometric relationships with each other, suggesting that they likely form different BRCA1 uniCOREs.

The interactions between UIMC1, FAM175A, FAM175B, and BRCC3/BRE/C19orf62 ("BRCA1-A" complex) have been described in the literature (Feng et al., 2009; Kim et al., 2007; Shao et al., 2009b; Wang et al., 2007, 2009a; Yan et al., 2007). In our dataset, FAM175A and FAM175B interact with UIMC1 in a mutually exclusive manner, and they resolve independently from each other in respective 3N analyses (Figures 2B and 2C). The FAM175A-UIMC1-BRCC3/BRE/C19orf62 uniCORE does not strongly bind known DNA-damage response components but coprecipitates CDC6/7 kinases, bridging into checkpoint BRCA1 functions (Figure 2D). In addition, we observed that this uniCORE preferentially interacts with the BRCA1/BARD1 MEMO, whereas the FAM175B-containing uniCORE can be found without UIMC1 and BRCA1 (Figure 2C). Specific functions of the latter complex, particularly in the BRCA1-independent context, are unknown.

Beyond UIMC1, BRCA1/BARD1 together with PALB2/BRCA2 bridge into mismatch repair processes via BRIP1 ("BRCA1-B" complex) and interactions with MLH1/PMS2 (Figure S2). BRCA1/BARD1/PALB2/BRCA2 also exists in a close CCI network with RBBP8 ("BRCA1-C" complex), the double-stranded break sensor MRN (*M*RE11-*R*AD50-*N*BS1) complex, CHEK1, RAD51, KEAP1, and MORF4L1/L2, implementing a BRCA DNA repair function. KEAP1, an E3 ubiquitin ligase adaptor (Cullinan et al., 2004; Kobayashi et al., 2004; Zhang et al., 2004), appears to be a component of the BRCA1 interactome; its function in this network remains to be investigated. Interestingly, the MORF4L1/2 MEMO itself has at least two
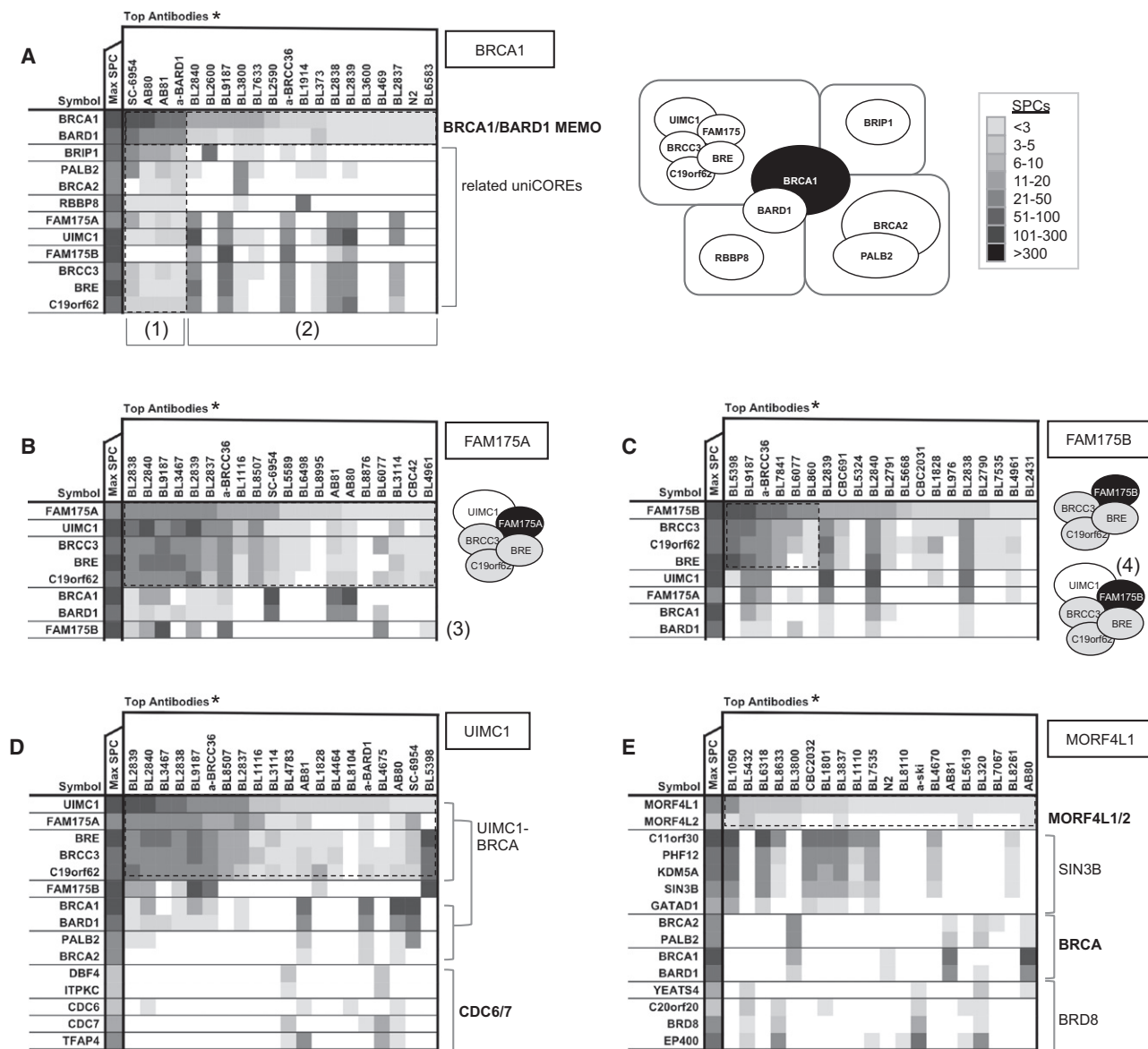
Figure 2. Protein Complex Heterogeneity in the BRCA1 Network

Partial BRCA1-related CCI networks, where individual MEMOs are separated by horizontal lines, are shown to highlight relationships between major components of the BRCA1 interactome (see also Figure S2). (*) Column headers specify antibody names, not the intended antigens that were used to generate the antibody. Despite conventions, it is often misleading to label IP/MS experiments with intended antigens because majority of antibodies cross-react with two or more proteins and some do not IP the intended antigen at all (see Table S1 for more information on cross-reacting antigens).

(A) Discernible "hierarchical" organization of protein interactions illustrated by a selection of IPs containing BRCA1 uniCOREs with the highest SPC identifications for all precipitated proteins (compare TopAntibodies lanes with MaxSPC lane). Although all proteins are equally true interactors of BRCA1, extensive reciprocal evidence visually implicates exclusive patterns: first three columns are BRCA1 and BARD1 IPs (1), where all BRCA1-containing complexes are shared, whereas experiments that target specific uniCOREs (2) show nonuniform distributions of BRCA1 interactors. To the right is a schematic interpretation of the findings.

(B) Top experiments for FAM175A reveal a stoichiometric complex between FAM175A, UIMC1, and BRE/BRCC36/C19orf62, but not with FAM175B (3).

3N analyses for FAM175B (C) and UIMC1 (D) show that all IPs that have FAM175A predominantly contain UIMC1 and BRCA1 (4) and thus reveal autonomous FAM175A/B uniCOREs.

(E) MORF4L1/2 MEMO interacts with BRCA1/BARD1/PALB2/BRCA2 uniCORE and also forms uniCOREs with chromatin-remodeling complexes SIN3B and BRD8.

additional uniCOREs with transcriptional coregulator complexes SIN3B and BRD8 (Figure 2E), highlighting widespread heterogeneity in protein complex organization.

The BRCA example illustrates that interactome modules are separable by data analysis in silico. Moreover, a hierarchical representation of CCIs via MEMOs and uniCOREs is an

appropriate way to illustrate these distinct, biologically relevant entities. As we continue to appreciate the modularity in protein complexes, it becomes increasingly sensible to catalog the divergence of cellular networks anchored in proteins shared by multiple uniCOREs. In retrospect, the BASC complex we reported 10 years ago is actually a merged representation of BRCA1 CCI networks (Wang et al., 2000) that now also include other components (Feng et al., 2009; Kim and Chen, 2008; Shao et al., 2009a, 2009b; Zhang et al., 2009). Thus, our HT-IP/MS study allows a "systems view" of the BRCA1 network, predicting various PPI dynamics and their vitality in biology. This is a testimony for the merit of a single-source HT-IP/MS effort, where uniform experimental conditions are applied and more consistent data are obtained for the analysis of context-dependent CCI networks.

Iterative Interrogation Reveals CCI Topology

The most effective way to mine our data is to iteratively search for proteins within a particular CCI network to reveal nuances in network organization. For example, when the largest RNA polymerase Pol II subunit (POLR2A) is used as the seed, six known Pol II subunits are found along with most subunits of the Mediator and Integrator complexes, RPAP2, GPN1/GPN3, SPEN, ZMYND8/ZNF687/ZNF592, and CDK9/CCNT1 and ELL elongation complexes (Figures 3A and 3B). Reciprocal searches using MED15, INTS7, or RPAP2 all preferentially return Pol II subunits, confirming these extensive interactions (Figure S3). Closer inspection of the data shows that the RPAP2 network lacks Mediator coactivator but has GPN1/GPN3 and Integrator. Therefore, RPAP2 likely resides in a subnetwork of Pol II where Integrator and Mediator separate. Analysis of the MED15 data shows that this protein is found in two exclusive complexes: it is a part of the large Mediator complex but also interacts separately with an E3 ligase, TRIM11, which binds and degrades MED15 (Ishikawa et al., 2006). Our data imply that TRIM11 sequesters the MED15 subunit away from the main body of the Mediator complex to form a relatively stable and discrete complex. Accordingly, we have singled out MED15 as a separate MEMO to account for this exclusive interaction pattern.

Next we use SPEN (*split en*ds homolog; a.k.a., SHARP) as an example of a single-coregulator MEMO at the crossroads of different transcriptional processes. SPEN CCI analysis reveals previously unappreciated protein interactions (Figure 3C). SPEN's associations with Mediator explain its presence in the Pol II network. This analysis also reveals the HDAC3/NCOR corepressor complex as a prominent interaction partner of SPEN. Finally, we recover SPEN interactions with a WTAP (*W*ilm's *t*umor 1 *a*ssociated *p*rotein) mRNA splicing complex, independently from HDAC3/NCOR and Mediator. Collectively, SPEN appears at the intersection of at least three seminal processes of eukaryotic transcription: chromatin-remodeling, transcriptional initiation, and mRNA processing. These SPEN interactions also explain how it can act as either a coactivator or a corepressor of transcription.

To our knowledge, this is the first attempt to comprehensively investigate and address how diverse patterns of associations in HT-IP/MS datasets translate into a tiered organization of complex-complex interactions on a proteome scale. Next, we use transcriptional coregulators to show how this resource can be used to better understand existing, and to explore unexpected, connections in transcriptional regulation.

Nuclear Receptor Transcriptional Regulation

The NURSA Consortium curates and maintains a tally of over 300 NR coregulators (Lanz et al., 2006) (http://www.nursa.org and Table S1). In our dataset, 128 NR coregulators were recovered as direct antigens (see Figure 1B for examples), while an additional 185 were identified as interacting proteins or in unintended cross-reacting complexes. We can confidently assign 170 MEMOs that contain 211 known NR coregulators and ~300 additional proteins that are likely coregulators. Furthermore, CCI networks of NR coregulators currently show >3700 unique gene products that participate in coregulation. This is a considerable expansion of the NURSA coregulator list, suggesting that many more proteins can participate in the regulation of NR-driven transcription than previously thought.

Deconvolution of the CCI Network of SRC-3 Reveals a Distinct Interactive Subset of NR Coregulators

We noticed that many NR coregulators, although recovered at high numbers in multiple reciprocal IPs, did not consistently form preferred interactions. Such promiscuous interaction patterns were observed most often for multifunctional coregulators that perform extremely diverse cellular functions. We therefore broadly classify transcriptional coregulators into two classes according to the type of protein complexes they form.

Type I coregulators exist in relatively stable multisubunit steady-state complexes and show little variation in their composition, thus conforming to our conventional image of a protein complex. Examples of type I complexes include well-known Mediator (Gu et al., 1999), nucleosome remodeling and deacetylase (NURD) complexes (Xue et al., 1998), CoREST complex (Hakimi et al., 2002), HDAC3/NCOR (Guenther et al., 2000), and SWI/SNF (BAF/P-BAF) (Kaeser et al., 2008; Wang et al., 1996), all of which we recovered fully in our IP/MS dataset, along with lesser known RANBP9, PELP1 (Figure 4A and Figure S4A), and the Z3 complexes described below.

The type II coregulators, on the other hand, do not have consistent steady-state stoichiometric partners and often resolve into one-component MEMOs (Figure 4B and Figure S4B). Such substoichiometric associations likely represent the hallmark behavior for many coregulators, which is to provide a quick response to different cellular signals through dynamic context-specific associations. The oncogenic SRC-3 (a.k.a., AIB1 or NCOA3) is a prime example of a type II coregulator.

The multifunctional transcriptional coregulator SRC-3 is a common target for cellular growth programs (Anzick et al., 1997; Kuang et al., 2004; Torres-Arzayus et al., 2004; Lonard et al., 2007; Yan et al., 2006), adipogenic and energy balance (Louet et al., 2006; Coste et al., 2008), and control of mRNA translation of proinflammatory cytokines (Yu et al., 2007). Such functional diversity may be provided by associated proteins, which bind SRC-3 proteins that are differentially modified post-translationally in response to the activation of different signaling pathways (Li et al., 2008; O'Malley et al., 2008; York et al., 2010). SRC-3 itself is a target of multiple cell signaling pathways (Font de Mora and Brown, 2000; Giannì et al., 2006; Wu et al., 2004;
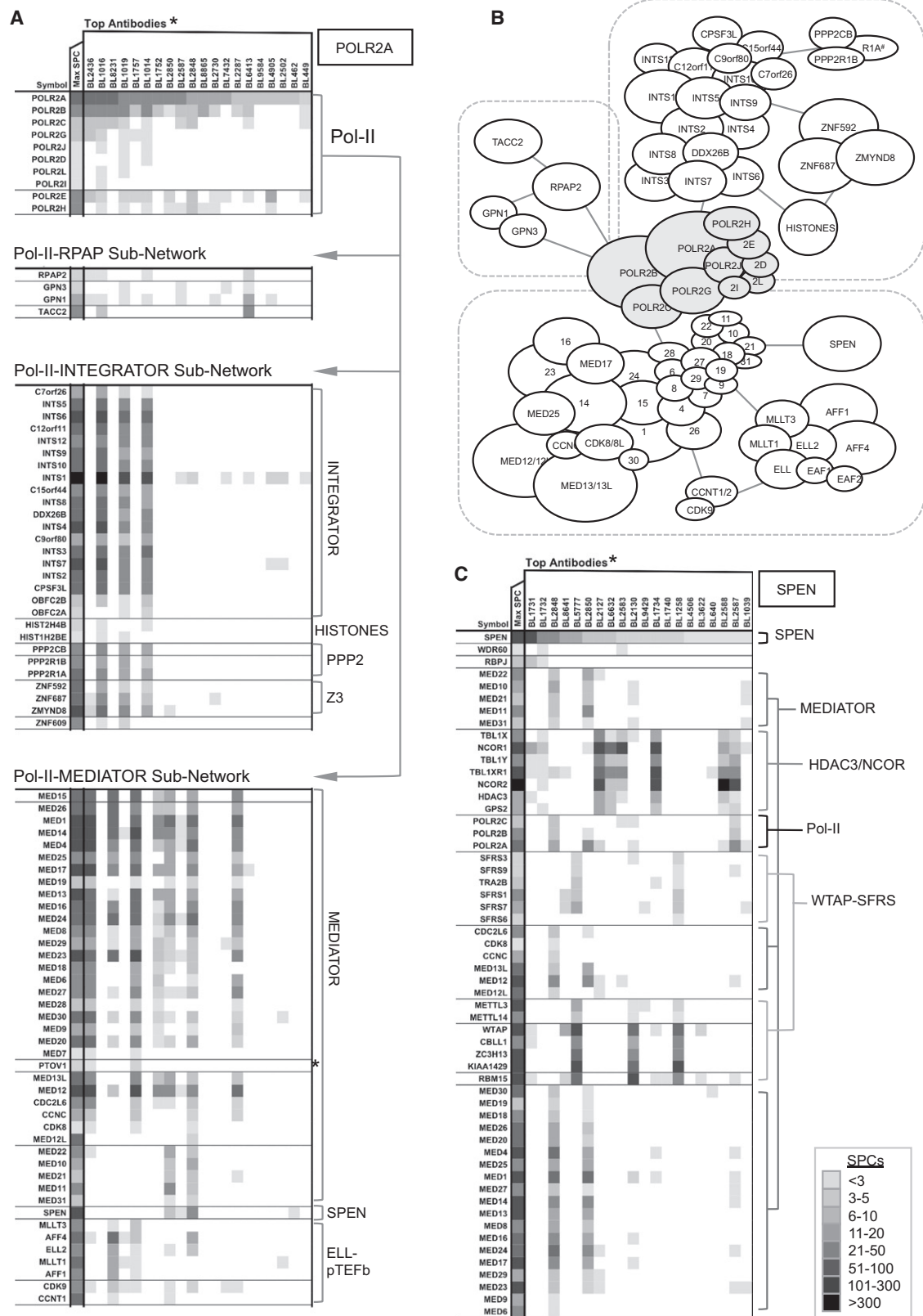
Yan et al., 2008; Yi et al., 2008) and acts as an integrator by converging signals to the genome for modulation of target gene expression. Our 3N analyses show only substoichiometric interaction patterns in the SRC-3 interactome, even for well-proven associations such as CBP/p300 and REGγ (PSME3) (Figure 4B). Along with the interacting coregulators, a diverse number of NRs, such as ERα, RXRα/β, and COUP-TFII, along with frequent tethering partners such as AP-1, coprecipitated with SRC-3 in our IP experiments. Importantly, this type II behavior is preserved in MCF-7 breast cancer cells described in our earlier study (Lanz et al., 2010). These observations suggest that type II behavior is an inherent and cell-type-independent property of this molecule, and quite possibly also of many other type II transcriptional coregulators.

The ability to distinguish type I from type II molecules serves as a good illustration of the conceptually different insights gained by a large-scale proteomics study, whose discovery-driven character is better positioned to reveal general perceptions, such as the overall biological functions of a type II protein. A classification of transcriptional coregulators based on their "interactivity" helped us to better understand why some coregulators have been reported to be responsible for many different cellular processes, whereas others seem to exert specialized biology only, albeit functioning in different signal and gene contexts. For type II molecules, variability of interactants found in individual experiments for such proteins is overcome by analyzing large single IP/MS studies like the one presented here. It then becomes clear that a protein that biochemically appears not to form a consistent complex often displays the most complicated regulatory interaction networks, indicating its important functional significance at the CCI level rather than in the formation of core complexes. The identification of type II interacting partners is indispensable for the understanding of regulatory processes and will remain an important challenge for affinity-based approaches and for translational efforts. This is certainly the case for oncogenic SRC-3.

The Z3 Coregulator Complex Participates with an Extensive CCI Network

Among hitherto unspecified coregulators, our work revealed a transcriptional coregulator complex whose number of interactions rivals that of Mediator or Integrator (Figure 5A). This Z3 MEMO consists of ZMYND8, ZNF687, and ZNF592. ZMYND8 contains a PHD-finger, a Bromodomain, a Pro-Trp-Trp-Pro (PWWP) motif, and a MYND-type zinc finger and is likely to function as a reader of histone modifications. It recently has been suggested to affect chromatin silencing (Poleshko et al., 2010) and to interact with the transcriptional coregulator RCOR2 (Zeng et al., 2010). ZNF687 and ZNF592 have multiple C2H2 zinc fingers known to bind DNA.

Z3 forms a uniCORE with a two-protein MEMO made of TSPYL1/TSPYL2 (testis-specific protein Y-like 1 and 2) (Figure 5A and Figure S5A). The Integrator protein complex appears to be the predominant partner of Z3, bringing Pol II into the Z3 CCI network. Z3 also is frequently found associated with the H3K4 demethylation machinery, including the H3K4me2/me1 demethylase KDM1 uniCORE, a partnering H3K4me3 demethylase KDM5A-containing SIN3B complex, and KDM5C H3K4me3 demethylase. Such conglomeration of histone modification readers suggests a prominent role for Z3 biology in interpreting the histone code for remodeling of chromatin for transcription. In addition, Z3 co-occurs with many transcription factors in our IP/MS dataset, including the nuclear receptor NR4A1 (NUR77, Figure S5B) and ERα, which we discuss next.

ERα interacts with the Z3 network at substoichiometric levels (Figure 5B) in MCF-7 breast cancer cells. We confirmed this by coimmunoprecipitation (co-IP) and in vitro GST pull-down assays (Figures 5C and 5D and Figure S5C). Furthermore, ChIP-seq analysis showed a significant overlap of ZMYND8 chromatin-binding sites with known ERα-binding sites (Figure 5E and Figure S5D). We also confirmed transcriptional coactivator functions of ZMYND8 in ERα-driven reporter assays and on endogenous E2-dependent genes (Figures 5F and 5G). siRNA knockdown of ZMYND8 showed markedly decreased transcription at the presumptive ERα/Z3 target genes ADORA1 and NAV2, whereas the classical ERα targets pS2/TFF1 and GREB1 appear to be less affected (Figure 5G), suggesting likely gene specificity of ZMYND8. This example again demonstrates the potential of our CCI resource to serve as a basis for deriving testable hypotheses.

Bridging the Gap between Genome-wide Data and Functional Proteomic Analyses

Currently, genome-wide association studies are identifying common genetic factors that impact health and disease. Even at an early stage the linking of genomics and proteomics data can further our understanding of disease pathways and provide platforms for generation of hypotheses. For example, in oncology, it is imperative to distinguish causative driver mutations from "hitchhiking" passengers (Stratton et al., 2009). Based on the premise that protein complexes function as biological units, we must imply that proteins that physically associate with known cancer drivers "collectively" confer a selective growth advantage on cells. These "guilt-by-association co-drivers" thus appear as additional valuable targets for functional translational analyses and therapeutic interventions. Cellular protein complexomes such as the resource presented here can complement genomic data with most valuable proteomics information, thus leading to a better understanding and treatments of human cancers.

Figure 3. Iterative Mining of HT-IP/MS Resource Reveals Topology of RNA Polymerase II Network

(A and B) 3N analyses for POLR2A show three separate subnetworks containing RPAP2-GPN1/3, Integrator complex, and Mediator complex (see also Figure S3 and Figure S6). The MEMOs in the POLR2A network were grouped by patterns of their distribution to show that Mediator and Integrator subnetworks are vividly independent of each other.

(C) Coregulator SPEN stands on crossroads of three seminal transcriptional processes. Its CCI network suggests an association with transcription through Mediator and chromatin remodeling via HDAC3/NCOR complex. SPEN also coprecipitates with proteins that regulate splicing process, including multisubunit SFRS and WTAP complexes. (*) Heatmap column headers specify antibody names, not the intended antigens that were used to generate the antibody (for antibody data, see Table S1).
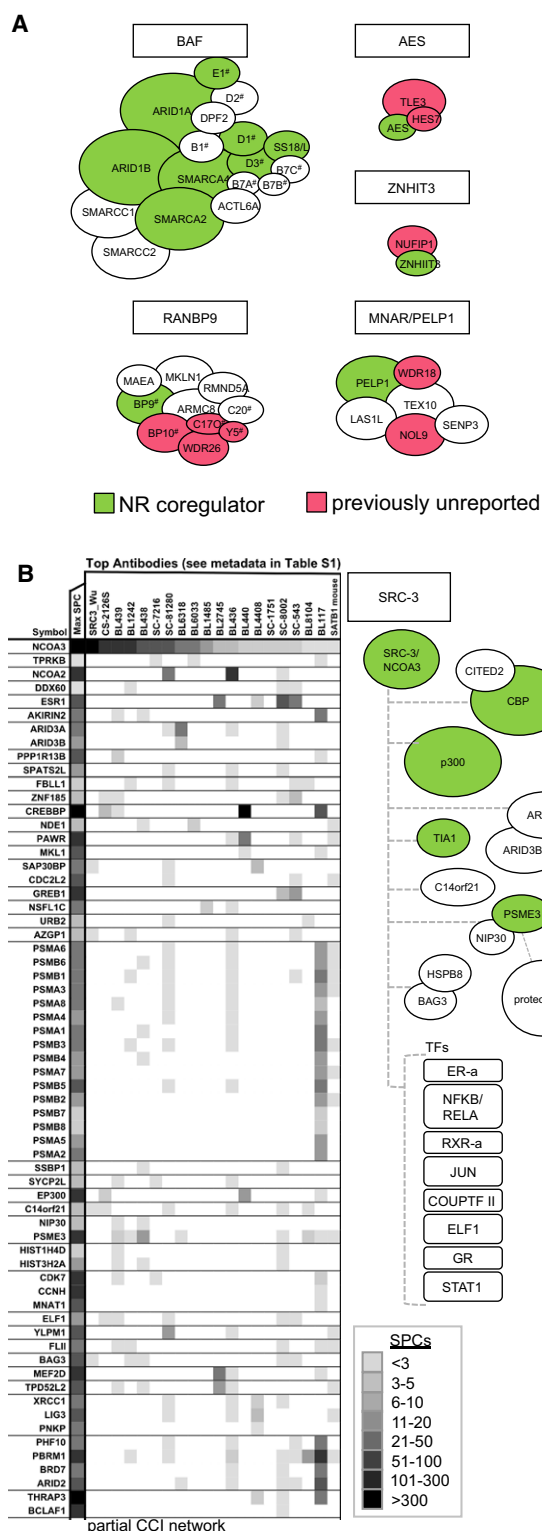
Figure 4. Coregulators Show Two Classes of Protein Interaction Profiles

3N analyses broadly classify coregulator networks into having either stable preferential protein networks (type I) or multiple transient interactions (type II) (heatmaps and additional examples are shown in Figure S4).

In Figure 6 we show two examples of the accrual of oncogenetic data in protein complexes. We used the Sanger Cancer Gene Census resource of human oncogenic mutations (Futreal et al., 2004) as well as the Broad Institute's Tumorscape database, which lists copy number changes across multiple cancer types (Beroukhim et al., 2010), to flag cancer-related gene products in our protein complex networks (see also Figure S6). In the P-BAF uniCORE, for example, an overwhelming number of complex components have different associations with cancers (Figure 6A), strongly suggesting a key physiological role of this chromatin-remodeling complex as a whole in oncogenesis, and implying that its subunits are functionally more likely to be drivers rather than passengers.

Similarly, the SIN3B uniCORE has three subunits that are significantly amplified in breast cancers (C11orf30, GATAD1, PHF12), whereas two more proximal components (KDM5A, ZNF131) are amplified in epithelial lineages (Figure 6B) (Beroukhim et al., 2010). Among these, C11orf30 (EMSY) and KDM5A were identified as drivers in breast cancers (Brown et al., 2010; Raouf et al., 2005; Santarius et al., 2010) and in acute myeloid leukemia (Futreal et al., 2004; Glaros et al., 2007; Wang et al., 2009b), respectively. Although the other gene products in this complex have not yet been causally implicated in cancer, our finding that they reside in a stable complex with validated cancer drivers predicts that they also are likely to provide a selective advantage to clonal cell expansions.

Perspectives

Here, we present a large-scale HT-IP/MS dataset and discuss examples of extensive interaction networks derived from endogenous human regulatory protein complexes. Our approach serves as an operational framework for the deconvolution of high-content proteomics data and testifies to the feasibility of mapping the endogenous protein complexome on a proteome-wide scale in higher eukaryotes.

The dense reciprocity of protein associations in our IP/MS dataset ensures high-confidence assignments and provides the highest standard of verification for native protein-protein interactions. The analysis of multiple reciprocal immunoprecipitations forgoes the necessity to exhaustively characterize the specificity of primary antibodies and produces a low number of false positive designations.

We recovered most published core complexes and have found many previously undocumented complex subunits. We also accomplished a characterization of weak protein interactions beyond most prior studies. As a result, we offer our hierarchical organization of endogenous protein complex interactions as a three-tiered organization of stable MEMOs, which combine

(A) Schematic illustrations of examples of type I NR coregulators with previously unidentified subunits. Green: coregulators listed at NURSA.org; E1, D2, D1, B1, and D3 are SMARC subunits; B7A, B, C are BCL7A, B, and C; C20, C20orf11; C17, C17orf39; BP9 and BP10 are RANBP9 and RANBP10; Y5, YPEL5, respectively.

(B) 3N heatmap excerpt for SRC-3/NCOA3. This is a typical type II coactivator that lacks stoichiometric steady-state complex while revealing a multitude of substoichiometric interactions, such as CBP (CREBBP), p300 (EP300), REG-gamma (PSME3), and various transcription factors (TFs).

to form variant core complexes (uniCOREs) that then interact with each other to form CCIs. Our resource includes snapshots of high-order associations between regulatory complexes and serves as an initial roadmap for understanding the human interactome.

Because endogenous protein complexes were isolated and identified, our method is directly transferable to any cell line in culture and will be extended to isolated tissues in the near future. This is particularly important when specificity for cell types and sensitivity to challenges such as hormonal treatments or exogenous stress must be considered to delineate particular context-dependent interactions. Ultimately, mapping the context-dependent interactomes is a molecular requisite for understanding cellular biology.

This study emphasized nuclear transcriptional coregulators and in type I and type II complexes revealed two broadly distinct coregulator interaction patterns. Further, we argue that when intelligently combined with genomics and transcriptomics data, proteomics information provides a rationale and predictable understanding of phenotype that has a real potential for translation. Our recurrent finding that cancer gene products group together in select protein complexes supports the idea that the perturbation of a protein complex as a whole should be implicated in the etiology of malignancies, entailing all complex constituents as valuable molecular targets for disease screening and therapy. This same logic can be applied to the etiology of polygenic metabolic or central nervous system diseases, where mutations in multiple genes are responsible for a disorder. We postulate that in many cases, polygenic mutations compromise the physiologic function within a single multiprotein complex, and then additional mutations further degrade the function of the particular interaction network above the threshold for disease progression. In this manner, our resource will provide ample ideas for further testing of the involvement of coregulatory proteins in the development of human polygenic diseases.

Taken together, genomics, proteomics, and targeted functional studies comprise three critical components of the combinatorial "systems" approach that can be employed to effectively translate genomics data into molecular knowledge in order to ultimately advance clinical applications.

EXPERIMENTAL PROCEDURES

Cell Culture and Nuclear Extraction

HeLa S3 were cultured in suspension in 20–36 l spin bottles in RPMI-1640 media with 5% FBS to final density of 0.5×10^6 cells/ml. Attached cell lines were grown in 50 to 200 15-cm plates in DMEM with 5% FBS. Cell fractionation was done as previously described (Malovannaya et al., 2010). For HeLa S3, nuclear extraction with 0.02–1.2M KCl buffer titration was applied; for MCF-7 cells, the 300–900 mM KCl gradient was used. Resulting ∼500 mM salt extract was dialyzed to 160 mM KCl.

Immunoprecipitation

IPs were carried out essentially as described (Malovannaya et al., 2010). Briefly, nuclear extracts were ultracentrifuged (200,000 rcf) after thawing; 0.5–1 ml extract (∼10–15 mg) was incubated with 7–15 μg primary antibody for 2 hr followed by ultracentrifugation and 45 min incubation with Sepharose-CL4B Protein A beads (GE Healthcare). Minimal bead washing with NTN (50 mM Tris-Cl, pH 8.0, 150 mM NaCl, 0.5% NP-40) was used to preserve tran-

sient interactions. Information about intended and predicted antigens of particular antibodies can be found in Table S1.

SDS-PAGE and Mass Spectrometry

Immunocomplexes were eluted in 1× Laemmli buffer and resolved on pre-cast 4%–20% Novex Tris-Glycine or 4%–12% NuPage gels (Invitrogen, CA, USA) to half-length. Gels were minimally stained with Coomassie brilliant blue, cut into 6 molecular weight ranges and heavy chain IgG band, and digested with trypsin. Immunocomplexes were identified on a Thermo Fisher LTQ (majority) or Velos-Orbitrap mass spectrometers. Spectral data were then searched against human protein RefSeq database in BioWorks or Proteome Discoverer Suites with either SeQuest (for LTQ data) or Mascot (Orbitrap data) software. Multi-consensus result files of protein GI identifiers were compiled for each IP with strict filters and often manually inspected.

Data Processing

IP/MS results were transferred into an in-house built FileMaker-based relational database where protein identification numbers (protein GIs) were converted to the GeneID identifiers according to the NCBI "gene2accession" table. Data filtering and deconvolution were performed as described in the text and in Extended Experimental Procedures.

Overexpression and Reciprocal co-IP

293T cells were transiently transfected with expression constructs for estrogen receptor (pCR3.1-hERα), pEGFP-C1-ZMYND8 (inserted at BglII/EcoRI sites), or control plasmids using Lipofectamine 2000 (Invitrogen). Forty-eight hours after transfection, cells were lysed in BC lysis buffer (20 mM Tris-HCl, pH 7.5, 0.2 mM EDTA, 20% glycerol, 0.15 M NaCl, 0.5% NP40, 10 mM β-ME, and protease inhibitors [PIs]), and lysates were centrifuged at 100,000 rcf and immunoprecipitated using 1 μg anti-ER (SC-8002, SCBT), anti-GFP (custom, Genemed Synthesis), or control rabbit/mouse IgG mix (Cell Signaling), followed by SDS-PAGE and western blotting.

RNA Interference and Real-Time Quantitative PCR

3×10^5 MCF-7 cells were transfected with 100 pmoles siRNA against ZMYND8 (HSS119059 and 1199061) or Negative Control Medium CG Content #2 siRNA (Invitrogen) using Lipofectamine RNAiMAX reverse transfection protocol. Cells were changed to 5% stripped serum DMEM media w/o phenol red at 16 hr and at 36 hr treated with 10 nM β-estradiol (Sigma) and collected 24 hr after treatment. Total RNA was isolated using RNeasy kit (QIAGEN); cDNA was prepared using SSIII RT system (Invitrogen) and analyzed by quantitative RT-PCR on ABI 7500 or StepOne Plus thermocyclers (Applied Biosystems) using SYBR Green (Applied Biosystems/LifeTechnologies) and gene-specific primers (see Figure S5D).

Luciferase Reporter Assays

HeLa cells were transfected with pCR3.1-hERα, pERE-E1b-LUC, pCMV-Tag2-SRC3, pCRIITOPO-ZMYND8, and/or control vectors using Lipofectamine 2000 (Invitrogen) and cultured in 5% stripped serum-containing DMEM w/o phenol red. Forty-eight hours after transfection, cells were treated overnight with 10 nM β-estradiol or ethanol vehicle, lysed in luciferase lysis buffer (25 mM Tris-HCl, pH8.0, 150 mM NaCl, 0.5% Triton X-100, 10% glycerol, 1 mM EDTA, 1 mM DTT, PIs), and luciferase was measured using Promega luciferase assay system on Centro LB960 luminometer (Berthold Technologies).

GST Pull-Down

GST fusion proteins (Figure S5C) were produced in BL21 bacteria (Stratagene) using 0.5 mM IPTG induction. Bacterial pellets were lysed in NETN buffer (50 mM Tris-HCl, pH8.0, 150 mM NaCl, 0.5% NP40, 5 mM EDTA, PIs), cleared by ultracentrifugation, and bound to glutathione sepharose 4B (GE Healthcare) for 1 hr at 4°C. Beads were washed with NETN, and amount of GST protein was estimated by SDS-PAGE followed by Coomassie staining. GST fusion proteins or control GST bound to beads were preincubated with 100 μl NETN containing 10 μg/ml BSA for 1 hr at 4°C to block nonspecific binding. Recombinant estrogen receptor was obtained from Invitrogen, diluted in NETN to 0.2 μg/ml, and cleared by ultracentrifugation. One hundred microliters
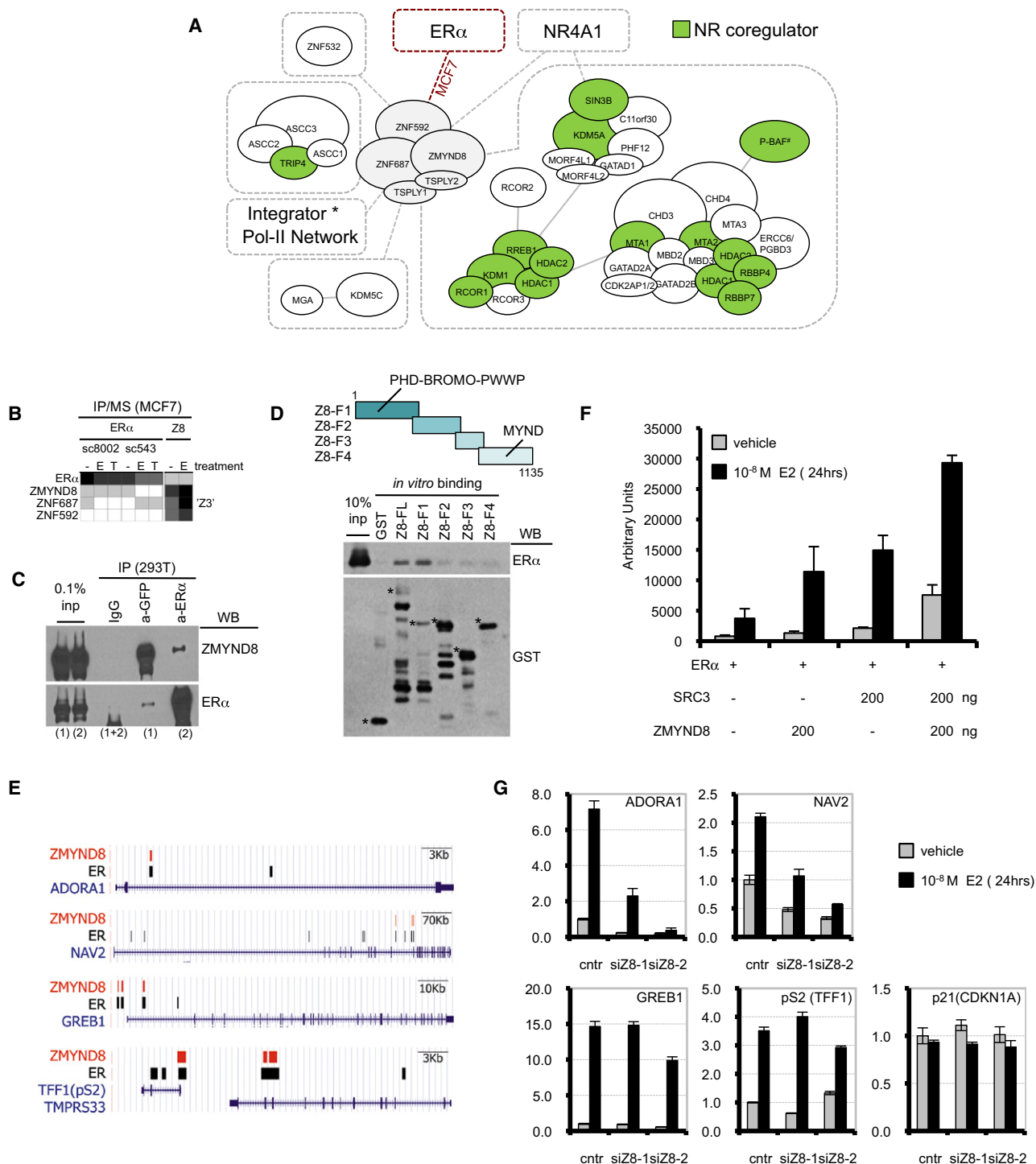
Figure 5. Z3 Complex Is a Transcriptional Coregulator

(A) Concise representation of the Z3 CCI network showing extensive interaction connections to transcriptional machinery (see also Figures S5A and S5B). Pol II-Integrator network (*) is omitted due to space limitations. In addition to histone demethylases KDM5C, KDM5A, and KDM1 (see text), Z3 also interacts with ASC1 (TRIP4) coregulator. (#) P-BAF, Polybromo and Brg/Brahma-Associated Factor (see Figure 4A for BAF). Known NR coregulators are shown in green.

(B) Reciprocal IP/MS of Z3 proteins with ERα in MCF-7 cells.

(C) Reciprocal IP/WB of overexpressed GFP-ZMYND8 and ERα in 293T cells; (1) and (2) are corresponding inputs.

(D) In vitro binding suggests that binding to ERα is via the N-terminal portion of ZMYND8 (Z8-F1). Asterisks indicate positions of ZMYND8 fragments.

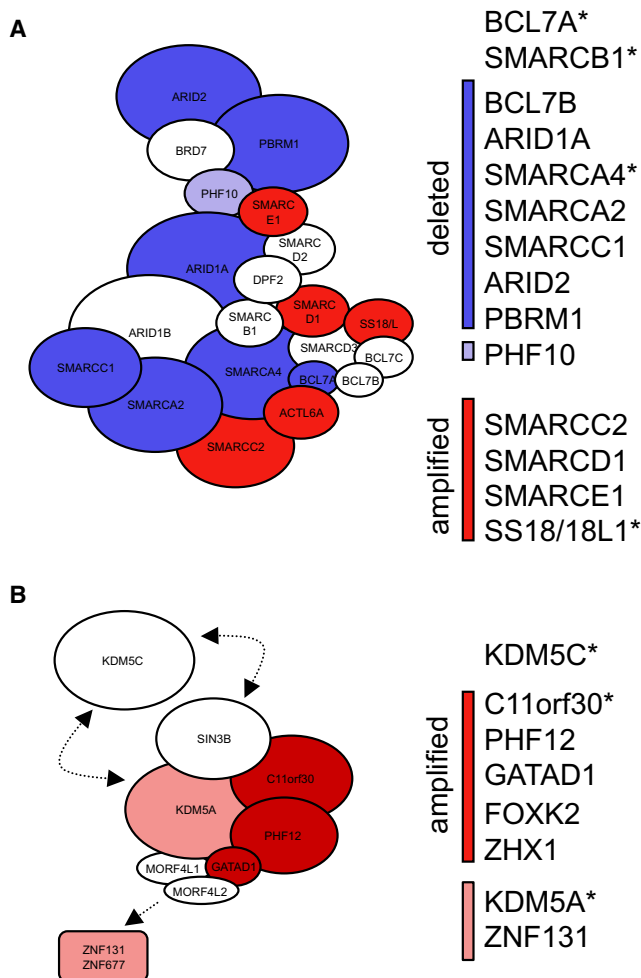(E) UCSC browser examples showing co-occupancy of ERα and ZMYND8-binding sites at E2-upregulated genes.

BCL7A*
SMARCB1*

deleted
BCL7B
ARID1A
SMARCA4*
SMARCA2
SMARCC1
ARID2
PBRM1
PHF10

amplified
SMARCC2
SMARCD1
SMARCE1
SS18/18L1*

KDM5C*

amplified
C11orf30*
PHF12
GATAD1
FOXK2
ZHX1

KDM5A*
ZNF131

Figure 6. Cancer Gene Alterations Group within Protein Complex Modules

(A) The P-BAF complex is significantly perturbed in lung cancers. Red: amplified; dark blue: deleted in non-small cell lung cancer (NSCLC); light blue: deleted in lung lineages.

(B) The SIN3B uniCORE is a hub of proteins with genomic amplifications. All but two proteins of the SIN3B complex are significantly amplified in breast cancer lineages (dark red) or epithelial lineages (pink). All genes marked with an asterisk are listed in Cancer Gene Sensus (Futreal et al., 2004) and/or have been implicated causally in cancer development. Amplification and deletions are marked according to Tumorscape resource (Beroukhim et al., 2010).

of ERα solution was added to 100 μl of bead-bound domains and incubated for 20 min at 4°C, followed by triple NETN washes. Pull-downs were analyzed by western blotting.

Genome-wide ZMYND8 Location Analysis by ChIP-Sequencing

Chromatin immunoprecipitation was performed according to the protocol from Upstate Biotechnology with minor modifications. 1×10^7 MCF-7 cells were used for each ChIP after 45 min treatment with 10 nM β-estradiol. After sonication the chromatin fraction was immunoprecipitated with CBC2032

ZMYND8 antibody and freed by heating. The ChIP-seq data were processed using MACS 1.3.5 with p values 1e8 and fold-change greater than 40 for peak calls.

REFERENCES

Alberts, B. (1998). The cell as a collection of protein machines: Preparing the next generation of molecular biologists. Cell 92, 291–294.

Anzick, S.L., Kononen, J., Walker, R.L., Azorsa, D.O., Tanner, M.M., Guan, X.Y., Sauter, G., Kallioniemi, O.P., Trent, J.M., and Meltzer, P.S. (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science 277, 965–968.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature 463, 899–905.

Brown, L.A., Johnson, K., Leung, S., Bismar, T.A., Benítez, J., Foulkes, W.D., and Huntsman, D.G. (2010). Co-amplification of CCND1 and EMSY is associated with an adverse outcome in ER-positive tamoxifen-treated breast cancers. Breast Cancer Res. Treat. 121, 347–354.

Cai, Y., Jin, J., Swanson, S.K., Cole, M.D., Choi, S.H., Florens, L., Washburn, M.P., Conaway, J.W., and Conaway, R.C. (2010). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J. Biol. Chem. 285, 4268–4272.

Coste, A., Louet, J.F., Lagouge, M., Lerin, C., Antal, M.C., Meziane, H., Schoonjans, K., Puigserver, P., O'Malley, B.W., and Auwerx, J. (2008). The genetic ablation of SRC-3 protects against obesity and improves insulin sensitivity by reducing the acetylation of PGC-1{alpha}. Proc. Natl. Acad. Sci. USA 105, 17187–17192.

Cullinan, S.B., Gordan, J.D., Jin, J., Harper, J.W., and Diehl, J.A. (2004). The Keap1-BTB protein is an adaptor that bridges Nrf2 to a Cul3-based E3 ligase: oxidative stress sensing by a Cul3-Keap1 ligase. Mol. Cell. Biol. 24, 8477–8486.

Feng, L., Huang, J., and Chen, J. (2009). MERIT40 facilitates BRCA1 localization and DNA damage repair. Genes Dev. 23, 719–728.

(F) ZMYND8 coactivates ERα in reporter luciferase assay; error bars represent standard deviation of three measurements.

(G) RNAi knockdown of ZMYND8 compromises upregulation of some E2-responsive target genes. p21 serves as a negative control; error bars represent standard deviation of three 18S-normalized mRNA measurements.

Font de Mora, J., and Brown, M. (2000). AIB1 is a conduit for kinase-mediated growth factor signaling to the estrogen receptor. Mol. Cell. Biol. *20*, 5041–5047.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Giannì, M., Parrella, E., Raska, I., Jr., Gaillard, E., Nigro, E.A., Gaudon, C., Garattini, E., and Rochette-Egly, C. (2006). P38MAPK-dependent phosphorylation and degradation of SRC-3/AIB1 and RARalpha-mediated transcription. EMBO J. *25*, 739–751.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., et al. (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727–1736.

Glaros, S., Cirrincione, G.M., Muchardt, C., Kleer, C.G., Michael, C.W., and Reisman, D. (2007). The reversible epigenetic silencing of BRM: implications for clinical targeted therapy. Oncogene *26*, 7058–7066.

Gu, W., Malik, S., Ito, M., Yuan, C.X., Fondell, J.D., Zhang, X., Martinez, E., Qin, J., and Roeder, R.G. (1999). A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. Mol. Cell *3*, 97–108.

Guenther, M.G., Lane, W.S., Fischle, W., Verdin, E., Lazar, M.A., and Shiekhattar, R. (2000). A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-repeat protein linked to deafness. Genes Dev. *14*, 1048–1057.

Hakimi, M.A., Bochar, D.A., Chenoweth, J., Lane, W.S., Mandel, G., and Shiekhattar, R. (2002). A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. Proc. Natl. Acad. Sci. USA *99*, 7420–7425.

Ishikawa, H., Tachikawa, H., Miura, Y., and Takahashi, N. (2006). TRIM11 binds to and destabilizes a key component of the activator-mediated cofactor complex (ARC105) through the ubiquitin-proteasome system. FEBS Lett. *580*, 4784–4792.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA *98*, 4569–4574.

Kaeser, M.D., Aslanian, A., Dong, M.Q., Yates, J.R., 3rd, and Emerson, B.M. (2008). BRD7, a novel PBAF-specific SWI/SNF subunit, is required for target gene activation and repression in embryonic stem cells. J. Biol. Chem. *283*, 32254–32263.

Kim, H., and Chen, J. (2008). New players in the BRCA1-mediated DNA damage responsive pathway. Mol. Cells *25*, 457–461.

Kim, H., Chen, J., and Yu, X. (2007). Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. Science *316*, 1202–1205.

Kobayashi, A., Kang, M.I., Okawa, H., Ohtsuji, M., Zenke, Y., Chiba, T., Igarashi, K., and Yamamoto, M. (2004). Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. Mol. Cell. Biol. *24*, 7130–7139.

Köcher, T., and Superti-Furga, G. (2007). Mass spectrometry-based functional proteomics: from molecular machines to protein networks. Nat. Methods *4*, 807–815.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Kuang, S.Q., Liao, L., Zhang, H., Lee, A.V., O'Malley, B.W., and Xu, J. (2004). AIB1/SRC-3 deficiency affects insulin-like growth factor I signaling pathway and suppresses v-Ha-ras-induced breast cancer initiation and progression in mice. Cancer Res. *64*, 1875–1885.

Lanz, R.B., Bulynko, Y., Malovannaya, A., Labhart, P., Wang, L., Li, W., Qin, J., Harper, M., and O'Malley, B.W. (2010). Global characterization of transcriptional impact of the SRC-3 coregulator. Mol. Endocrinol. *24*, 859–872.

Lanz, R.B., Jericevic, Z., Zuercher, W.J., Watkins, C., Steffen, D.L., Margolis, R., and McKenna, N.J. (2006). Nuclear Receptor Signaling Atlas (http://www.nursa.org): hyperlinking the nuclear receptor signaling community. Nucleic Acids Res *34*, D221-D226.

Li, C., Liang, Y.Y., Feng, X.H., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (2008). Essential phosphatases and a phospho-degron are critical for regulation of SRC-3/AIB1 coactivator function and turnover. Mol. Cell *31*, 835–849.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. (2004). A map of the interactome network of the metazoan C. elegans. Science *303*, 540–543.

Lonard, D.M., and O'Malley, B.W. (2007). Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation. Mol. Cell *27*, 691–700.

Lonard, D.M., Lanz, R.B., and O'Malley, B.W. (2007). Nuclear receptor coregulators and human disease. Endocr. Rev. *28*, 575–587.

Louet, J.F., Coste, A., Amazit, L., Tannour-Louet, M., Wu, R.C., Tsai, S.Y., Tsai, M.J., Auwerx, J., and O'Malley, B.W. (2006). Oncogenic steroid receptor coactivator-3 is a key regulator of the white adipogenic program. Proc. Natl. Acad. Sci. USA *103*, 17868–17873.

Malovannaya, A., Li, Y., Bulynko, Y., Jung, S.Y., Wang, Y., Lanz, R.B., O'Malley, B.W., and Qin, J. (2010). Streamlined analysis schema for high-throughput identification of endogenous protein complexes. Proc. Natl. Acad. Sci. USA *107*, 2431–2436.

McKenna, N.J., Lanz, R.B., and O'Malley, B.W. (1999). Nuclear receptor coregulators: cellular and molecular biology. Endocr. Rev. *20*, 321–344.

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., et al. (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. Mol. Cell *21*, 811–823.

O'Malley, B.W., Qin, J., and Lanz, R.B. (2008). Cracking the coregulator codes. Curr. Opin. Cell Biol. *20*, 310–315.

Poleshko, A., Einarson, M.B., Shalginskikh, N., Zhang, R., Adams, P.D., Skalka, A.M., and Katz, R.A. (2010). Identification of a functional network of human epigenetic silencing factors. J. Biol. Chem. *285*, 422–433.

Raouf, A., Brown, L., Vrcelj, N., To, K., Kwok, W., Huntsman, D., and Eaves, C.J. (2005). Genomic instability of human mammary epithelial cells overexpressing a truncated form of EMSY. J. Natl. Cancer Inst. *97*, 1302–1306.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes—2009. Nucleic Acids Res. *38*(Database issue), D497–D501.

Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., and Cooper, C.S. (2010). A census of amplified and overexpressed human cancer genes. Nat. Rev. Cancer *10*, 59–64.

Shao, G., Lilli, D.R., Patterson-Fortin, J., Coleman, K.A., Morrissey, D.E., and Greenberg, R.A. (2009a). The Rap80-BRCC36 de-ubiquitinating enzyme complex antagonizes RNF8-Ubc13-dependent ubiquitination events at DNA double strand breaks. Proc. Natl. Acad. Sci. USA *106*, 3166–3171.

Shao, G., Patterson-Fortin, J., Messick, T.E., Feng, D., Shanbhag, N., Wang, Y., and Greenberg, R.A. (2009b). MERIT40 controls BRCA1-Rap80 complex integrity and recruitment to DNA double-strand breaks. Genes Dev. *23*, 740–754.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719–724.

Torres-Arzayus, M.I., Font de Mora, J., Yuan, J., Vazquez, F., Bronson, R., Rue, M., Sellers, W.R., and Brown, M. (2004). High tumor incidence and activation of the PI3K/AKT pathway in transgenic mice define AIB1 as an oncogene. Cancer Cell *6*, 263–274.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

Wang, B., Hurov, K., Hofmann, K., and Elledge, S.J. (2009a). NBA1, a new player in the Brca1 A complex, is required for DNA damage resistance and checkpoint control. Genes Dev. *23*, 729–739.

Wang, B., Matsuoka, S., Ballif, B.A., Zhang, D., Smogorzewska, A., Gygi, S.P., and Elledge, S.J. (2007). Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. Science 316, 1194–1198.

Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J., and Allis, C.D. (2009b). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature 459, 847–851.

Wang, W., Côté, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R., Muchardt, C., Kalpana, G.V., Goff, S.P., Yaniv, M., et al. (1996). Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. EMBO J. 15, 5370–5382.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S.J., and Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev. 14, 927–939.

Weake, V.M., and Workman, J.L. (2010). Inducible gene expression: diverse regulatory mechanisms. Nat. Rev. Genet. 11, 426–437.

Wu, R.C., Qin, J., Yi, P., Wong, J., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (2004). Selective phosphorylations of the SRC-3/AIB1 coactivator integrate genomic reponses to multiple cellular signaling pathways. Mol. Cell 15, 937–949.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol. Cell 2, 851–861.

Yan, J., Tsai, S.Y., and Tsai, M.J. (2006). SRC-3/AIB1: transcriptional coactivator in oncogenesis. Acta Pharmacol. Sin. 27, 387–394.

Yan, J., Kim, Y.S., Yang, X.P., Li, L.P., Liao, G., Xia, F., and Jetten, A.M. (2007). The ubiquitin-interacting motif containing protein RAP80 interacts with BRCA1 and functions in DNA damage repair response. Cancer Res. 67, 6647–6656.

Yan, J., Erdem, H., Li, R., Cai, Y., Ayala, G., Ittmann, M., Yu-Lee, L.Y., Tsai, S.Y., and Tsai, M.J. (2008). Steroid receptor coactivator-3/AIB1 promotes cell migration and invasiveness through focal adhesion turnover and matrix metalloproteinase expression. Cancer Res. 68, 5460–5468.

Yi, P., Feng, Q., Amazit, L., Lonard, D.M., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (2008). Atypical protein kinase C regulates dual pathways for degradation of the oncogenic coactivator SRC-3/AIB1. Mol. Cell 29, 465–476.

York, B., Yu, C., Sagen, J.V., Liu, Z., Nikolai, B.C., Wu, R.C., Finegold, M., Xu, J., and O'Malley, B.W. (2010). Reprogramming the posttranslational code of SRC-3 confers a switch in mammalian systems biology. Proc. Natl. Acad. Sci. USA 107, 11122–11127.

Yu, C., York, B., Wang, S., Feng, Q., Xu, J., and O'Malley, B.W. (2007). An essential function of the SRC-3 coactivator in suppression of cytokine mRNA translation and inflammatory response. Mol. Cell 25, 765–778.

Zeng, W., Kong, Q., Li, C., and Mao, B. (2010). Xenopus RCOR2 (REST corepressor 2) interacts with ZMYND8, which is involved in neural differentiation. Biochem. Biophys. Res. Commun. 394, 1024–1029.

Zhang, D.D., Lo, S.C., Cross, J.V., Templeton, D.J., and Hannink, M. (2004). Keap1 is a redox-regulated substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex. Mol. Cell. Biol. 24, 10941–10953.

Zhang, F., Fan, Q., Ren, K., and Andreassen, P.R. (2009). PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. Mol. Cancer Res. 7, 1110–1118.